https://doi.org/10.48417/technolang.2026.03.07
Research article

Between Determinism and Chance: From Llull's Machine to Generative AI Models

Daria Bylieva (✉)
Peter the Great St.Petersburg Polytechnic University, Polytechnicheskaya, 29, St. Petersburg, 195251, Russia
bylieva_ds@spbstu.ru

Abstract

In 1937, Jorge Luis Borges looked back some 650 years at Ramon Llull's machinic *ars inveniendi*. This is one of five essays that looks at contemporary learning machines through the lens of Borges's characteristically engimatic historical note and literary reflection. It examines the epistemological tension between determinism and randomness in projects of formalising thought – from Ramon Llull's *Ars Magna* (1305) to contemporary generative transformers. Borges's 1937 critique of Llull revealed a fundamental impasse: closed combinatorial systems, lacking contextual sensitivity, produce syntactic noise rather than truth. Yet Borges proposed an alternative, namely, to use such machines as generators of random combinations subject to human selection. This shift finds unexpected validation in AI transformer models. By replacing discrete symbols with continuous semantic embeddings and dynamic attention mechanisms, transformers overcome Llull's linear blindness. However, in deterministic mode (low temperature), they merely reproduce linguistic clichés, exhibiting neural text degeneration. Genuine novelty emerges only through stochastic sampling–temperature-based deviation into the probability distribution's long tail, actualising latent semantic projections (as in the Borgesian "red tiger"). Drawing on recent empirical studies, the paper demonstrates a fundamental trade-off between alignment due to reinforcement learning (RLHF) and stochastic creativity. As an alternative, calibrated uncertainty is proposed, allowing models to acknowledge the limits of knowledge. The transformer that legitimises its stochastic nature thereby approaches Borges's ideal of a poetic machine – an instrument that does not prove but suggests, that does not close off truth but opens a space for play.

Keywords: Ramon Llull; Jorge Luis Borges; Generative AI; Transformers; Temperature sampling; Randomness; Creativity; Combinatorics; Calibrated uncertainty; TRIZ

Citation: Bylieva D. (2026). Between Determinism and Chance: From Llull's Machine to Generative AI. Models *Technology and Language, 7*(3), 94-107.https://doi.org/10.48417/technolang.2026.03.07

© Bylieva D. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

УДК 7+117
https://doi.org/10.48417/technolang.2026.03.07
Научная статья

Между детерминизмом и случайностью: от машины Луллия к генеративным моделям ИИ

Дарья Сергеевна Быльева (✉)
Санкт-Петербургский политехнический университет Петра Великого (СПбПУ), Политехническая, 29, Санкт-Петербург, 195251, Россия
bylieva_ds@spbstu.ru

Аннотация
В статье рассматривается эпистемологическое напряжение между детерминизмом и случайностью в проектах формализации мышления – от Ars Magna Раймунда Луллия (1305) до современных генеративных трансформеров. Борхесовская критика Луллия (1937) вскрыла фундаментальный тупик замкнутых комбинаторных систем: лишённые контекстуальной чувствительности, они продуцируют синтаксический шум вместо истины. Однако Борхес предложил альтернативу – использовать такие машины как генераторы случайных комбинаций, подлежащих человеческой селекции. Этот сдвиг находит неожиданное подтверждение в архитектуре трансформеров: замена дискретных символов непрерывными семантическими эмбеддингами и динамическими механизмами внимания позволяет преодолеть луллианскую линейную слепоту. Однако в детерминированном режиме (низкая температура) трансформеры лишь воспроизводят языковые клише, демонстрируя нейронную дегенерацию текста. Подлинная новизна возникает только через стохастическое сэмплирование – температурную девиацию в “длинный хвост” распределения вероятностей, актуализирующую латентные семантические проекции (как в борхесовском “красном тигре”). Опираясь на новейшие эмпирические исследования, статья демонстрирует фундаментальный компромисс между выравниванием (RLHF) и стохастической креативностью. В качестве альтернативы предлагается калиброванная неопределённость, позволяющая моделям признавать границы собственного знания. Трансформер, легитимизирующий свою стохастическую природу, тем самым приближается к борхесовскому идеалу поэтической машины – инструмента, который не доказывает, но предлагает; не замыкает истину, но открывает пространство для игры.

Ключевые слова: Раймунд Луллий; Хорхе Луис Борхес; генеративный ИИ; трансформеры; температурное сэмплирование; случайность; креативность; комбинаторика; калиброванная неопределённость; ТРИЗ

Для цитирования: Bylieva D. Between Determinism and Chance: From Llull's Machine to Generative AI // Technology and Language. 2026. № 7(3). P. 94-107. https://doi.org/10.48417/technolang.2026.03.07

© Быльева Д. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

INTRODUCTION

For centuries, the rationalist tradition advanced under the assumption that thinking could be wholly purged of human subjectivity, ambiguity, and risk. From medieval ontological calculi to the early modern projects of a universal symbolic language, the prevailing conviction was that if the world is logically ordered and language consists of a finite set of elements, then the generation of new knowledge could be algorithmised through systematic recombination of signs. The dream of an infallible automaton rested on the dictum of absolute determinism, where the correct rotation of conceptual gears would guarantee the automatic extraction of truth.

The large-scale deployment of contemporary generative systems of distributed semantics has suddenly returned this centuries-old ambition to the epicentre of a profound epistemological crisis, exposing the immanent dead ends of blind syntax. Engineering ambitions that sought to construct an ideal semiotic machine, are confronted with fundamental anomalies of code: neural text degeneration and the phenomenon of confabulations (hallucinations). Under rigid probability maximisation, algorithms collapse into an autoregressive loop, endlessly producing banal clichés, tautologies, and dead templates. Conversely, the attempt to overcome this predictability unleashes the destructive nature of linear syntax – the machine begins to "bluff," masking the complete absence of a semantic criterion of truth with peremptory grammatical confidence. What is today called "bugs" of AI is analogously a systemic property of any closed combinatorics.

In 1937, Jorge Luis Borges published an essay devoted to the "thinking machine" of Ramon Llull – a medieval philosopher who, three centuries before Leibniz and seven centuries before the advent of computers, devised a mechanical combinatorial apparatus for proving divine truths. Borges identified a fork between the dead end of combinatorics for achieving truth and the possibility of creative breakthroughs on its basis. For a long time, this Borgesian shift was perceived exclusively as a metaphor; however, recent research in automated scientific discovery directly resurrects Llullian logic. In the work *The Ramon Llull's Thinking Machine for Automated Ideation*, researchers de facto reproduced the concentric disks of the *Ars Magna*, distributing contemporary scientific discourse along three rigid axes of combinatorial synthesis: Theme, Domain, and Method (Zhao et al., 2025). The experiment showed that forced, hard-wired crossing of elements along these axes allows overcoming the limitations of standard generation; yet the authors explicitly point to a fundamental lack of diversity and the inability of models to independently discover semantic novelty without an external stochastic impulse. The contemporary theory of large language models confronts the same contradiction as Rationalist scholasticism: how can a closed sign system breach its own boundaries without sliding either into a loop of degeneration or into the chaos of uncontrolled confabulations?

THE COMBINATORIAL LOGIC OF THE MACHINE

In his treatise *Ars Magna* (1305), Ramon Llull posited that being consists of a finite number of indivisible attributes – nineteen (in different versions) divine perfections, such

as goodness, greatness, eternity, power, wisdom, will, and others. These attributes were conceived as an alphabet whose mechanical combination would exhaust all truth. Llull believed in the isomorphism of language and reality: the order of signs must exactly correspond to the order of being. His machine consisted of concentric disks whose rotation generated all possible pairwise combinations of basic concepts.

The ideas described in Llull's *Ars Magna* found direct continuation in the works of Giordano Bruno and Gottfried Wilhelm Leibniz. The latter sought to expand the concepts used, creating a universal set of notions from which all others could be constructed. In early 1666, Leibniz wrote his treatise *De Arte Combinatoria*, proposing a project of a *characteristica universalis* – an alphabet of human thought. Leibniz argued that just as words consist of letters, complex truths are combinations of a relatively small number of simple concepts, which can be decomposed into simple ideas. Leibniz's crucial conceptual shift lay in translating logical inference into the plane of calculus: the rules of inference themselves became formal operations over symbols, which presupposed the possibility of resolving any theoretical dispute by the universal imperative *calculemus* ("let us calculate").

The idea, originating with Llull, of creativity as recombination of ready-made elements was reproduced in culture for centuries, revealing the same dead end of syntactic blindness. In mannerist and baroque poetry, this collage method was applied both in centos from fragments of Virgil (from abstract meanings to obscene ones in Ausonius) and in Christian adaptations such as *Virgilius evangelizans*, where the deterministic reassembly of fragments was meant to reveal Christian Truth. Later, this automatism crystallised in versification machines made of concentric circles, as in the *Cynosura* of Nicolaus of Lowitz (1653). Musical thought moved in parallel: from Athanasius Kircher's *Arca musarithmica* (1650), which allowed mechanical composition of hymns according to tables (Kircher, 1650), to the dice games of Johann Philipp Kirnberger, Maximilian Stadler, and *Mozart's Musikalisches Würfelspiel* [Mozart's Musical Game of Dice]. However, these systems merely shuffled ready-made clichés, without implying any semantic shift.

The idea of combinatorics of possible ideas as a means of attaining new knowledge was repeatedly satirised by various authors. Criticism was directed at the fact that mechanical rules operate with blind symbols and do not guarantee coherence. Early in the 18th century Charles Stanhope attempted to overcome these limitations and translate logical operations into a practical planeby creating devices for solving syllogisms with numerical quantifiers.

An important step toward the pragmatisation of combinatorial synthesis and the transition from abstract philosophical systems to work with applied data structures was the invention of Semen Korsakov. In 1832, he created mechanical prototypes of five "intellectual machines" using perforated cards (Mikhailov, 2016). Here combinatorics was narrowed to algorithms for a rigid search for correspondences. Concerning one such machine – the homeoscope – Korsakov wrote:

> Its application in the field of medicine will bring enormous benefit, since in the case of illness, based on a detailed enumeration of all symptoms, it (the device)

> can display with the highest degree of accuracy the medicine most suitable for that case, according to the desired medical methodology. (Korsakov, 1832/2009).

In general, the author noted the practical usefulness of machines in intellectual activity, indicating that they could be used in solving various problems in everyday life to make any kind of inference (Korsakov, 1832/2009, p. 9).

In this sense, the description of Babbage's Analytical Engine – considered to be the technological prototype of the modern computer – pursued much more modest and specialised goals. Charles Babbage designed an apparatus capable of performing any mathematical computation. Yet Ada Lovelace (1843), who wrote the first program for the Analytical Engine, saw in it not merely a calculator but a "thinking or ... reasoning machine" (p. 675). She made a fundamental semiotic breakthrough, suggesting that operations could be performed not only on numbers but also on other objects, provided their rules of interaction could be formalised:

> The essence and purpose of the machine will change depending on what information we put into it. The machine will be able to write music, paint pictures, and show science paths that we have never seen anywhere. (Lovelace, 1843, p. 694)

RANDOMNESS AS A SOURCE OF INSPIRATION

In his essay *The Thinking Machine of Ramon Llull* Borges criticised Llull's project on two levels. First, the combinatorial power of a finite alphabet leads to exponential explosion: already with nine categories, the number of pairwise combinations reaches thirty-six, and with longer chains grows factorially. Yet the explosion is merely a quantitative problem. Far more significant is the qualitative failure: the mechanism does not distinguish meaningful combinations from meaningless ones. For the machine, the statement "God is eternal goodness" and an arbitrary set of signs have the same status, since Llull's syntax is linear and blind to context. Semantic distance between concepts is not computed; connections are rigidly fixed by the rotation of disks, and no feedback allows for the filtering out of noise. Borges writes that the machine "works too well" – it gives an answer to any question, but none of these answers is accompanied by a criterion of truth. Nevertheless, Borges makes an unexpected concession: if we abandon the pretension to logical proof and consider the machine as a generator of random combinations, its exhaustive enumeration can serve as raw material for poetic imagination. He shows that the inner mechanism of the author's work on a text is structurally identical to the combinatorial automaton:

> The poet in search of an epithet acts in much the same way as the machine. He tries out different epithets until he finds one that is sufficiently striking. 'Black tiger' might do for a tiger in the night; 'red tiger' – because of its association with blood – for all tigers. (Borges, 1999, p. 159)

The poet, like the Llullian machine, operates with a finite set of elements (epithets), combining them with the object (tiger) until a combination emerges that is "sufficiently

striking" – that is, one that violates expectations while preserving coherence. This analogy clarifies what Borges means by the "astonishing" shift. It is not randomness as such, but a selective act: the machine (or the poet) produces many combinations, but only one of them stops the enumeration by being recognised as valuable. It is precisely this act of selection, according to Borges, that turns combinatorial noise into poetry. As Eduardo Guízar (2012) notes in analysing this passage, Borges "advises applying the machine to poetry, but not before recognising that it is a 'ridiculous machine,' an 'illusory machine' " (p. 175).

For Borges himself, this mechanics of shifting and eluding meaning is closely linked to his main personal metaphor – the image of the tiger. The writer confessed:

> The first color I remember is the yellow of the tiger's skin; and to this day, having lost my sight, I continue to see yellow, among a few other colors. (Borges, 1984, p. 71).

But the tiger in his world paradoxically escapes: it always remains unattainable, existing only in language and imagination. In the miniature *Dreamtigers*, Borges describes how he had dreamed of these predators since childhood, but whenever he tried to consciously and deterministically summon the image of a tiger in his dream, the control mechanism broke down: the dream would present him with birds or shape-shifting monsters. Rigid, logical prescription cannot generate a living image.

In the poem *The Other Tiger* (*El otro tigre*) Borges (1960/1985) contrasts the real tiger that roams by the Ganges with the tiger of "symbols" born of verse. The poet knows that his tiger is merely a "shadow" and a "creation of words," yet refuses to cease his futile pursuit. This symbolic reduction aligns perfectly with the mechanics of Ramon Lull's logical machine, where the living beast is deconstructed into a pure semiotic sign to fit a combinatorial model of the cosmos.

In *The Book of Imaginary Beings* this Lullian approach reaches its peak within the myth of the Tigers of Annam. Here, the traditional heterogeneous pantheon of Chinese guardians is replaced by a homogeneous tiger monad multiplied into five spectral and elemental modes. The coordinates of this poetic geometry are strictly fixed: the Red Tiger commands the South (summer, fire) at the top of the map, mirrored by the Black Tiger in the North (winter, water). The horizontal axis is held by the Blue Tiger in the East (spring, plants) and the White Tiger in the West (autumn, metals), while the sovereign Yellow Tiger anchors the metaphysical Centre. Through this systematic permutation of colors and elements, Borges transforms the elusive archetype of the tiger into a universal combinatorial alphabet capable of structuring the entire universe. Yet, this rigid Lullian geometry, which seeks to contain reality through precise categorization, inevitably collapses when confronted with the uncanny (Borges, 1967/2005).

The ultimate failure of any rational or combinatorial matrix to map the infinite is fully realized in his late story *The Blue Tigers*, where the protagonist finds strange blue stones that multiply chaotically and disappear when one tries to count them. These blue stones are a metaphor for pure chance that breaks the inertia of reason. If the Tigers of Annam represent the pinnacle of symbolic order – a cosmos where even the wild beast is

tamed into a predictable coordinate – the blue stones manifest its absolute subversion. They represent a chaotic counter-weight to Lull’s machine: an unquantifiable, shifting plurality that defies arithmetic and resists being trapped in any verse or cosmological system.

The principle of using randomness as a trigger for generating semantic novelty can be found in the works of many artists and writers. Nikolai Gogol describes Alexander Pushkin's method: the poet wrote headings on paper labels – “Russian Winter,” “Derzhavin,” names of objects – put them into a vase, and then “drew out the first label at random,” after which he wrote down everything that came to mind in connection with that random name (Gogol, 1844). This is not merely a fixation of associations but a controlled stochastic procedure: an external random stimulus serves as a trigger for retrieving implicit connections that would otherwise remain unactualised. Similarly, according to Yuri Lyubimov, Sergei Yesenin used a box of words, scattered them around the room, and then “crawled around and composed poems from them” – a procedure where the physical randomness of spatial distribution became a source of syntactic novelty (Lyubimov, 2001, p. 402). Velimir Khlebnikov cut his texts into parts and shuffled them to break the ossified structure of language (Vroon, 1983, pp. 24–27). Later, in the twentieth century, William Burroughs and Brion Gysin formalised this approach. By cutting printed pages into four parts and randomly rearranging them, they found that at the joints of alien phrases, prophetic, metaphorical meanings emerged that were inaccessible to reason. Burroughs claimed that “cut-ups” reveal the hidden code of reality (Burroughs & Gysin, 1978, pp. 29–33).

In the visual arts, this stochastic impulse can be traced from antiquity – when, according to Pliny the Elder (c. 77 CE/1991), the painter Protogenes found inspiration in a chance mark left by a sponge thrown at his painting – to Leonardo da Vinci, who in his 1651 treatise recommended drawing ideas from random stains on “painted walls” (Ugander & Epstein, 2024). In the twentieth century, the avant-garde turned this chaos into method. Tristan Tzara created Dadaist poetry by pulling words from a hat, Marcel Duchamp in “3 Standard Stoppages” (1913–1914) fixed the lines of fallen threads, and John Cage used the ancient Chinese *I Ching* for composing aleatoric music. The culmination of this approach came with Jackson Pollock's “controlled loss of control” in drip painting: allowing gravity and viscosity to form the lines, he constantly adjusted paint density, turning random events into a deliberate structure endowed with powerful authenticity.

However, randomness can serve as a stimulus not only for creating literary or artistic objects but also in other fields, notably invention (Demina & Nesterov, 2025). The key problem that the Russian Theory of Inventive Problem Solving (TRIZ) places at the centre of its epistemology is a psychological inertia of thinking – a fundamental cognitive mechanism by which thinking spontaneously narrows down the search space to include only established, tried and true solutions, thus excluding entire classes of possibilities even before any analysis (Kowalick, 1998). To counteract this inertia, various methods of creative activation rely on randomness, operationalizing it as controlled semantic noise: the forced introduction of an element not logically connected to the current task to stimulate associative networks.

A classic example of such a mechanism is the Method of Focal Objects (MFO), which Altshuller (1996) analyzed as a tool for breaking cognitive biases by transferring the features of several randomly selected objects onto the object being improved. The technical procedure is as follows: an object to be modified (the focal object) is chosen; then, several random objects are selected (e.g., by opening a dictionary to a random page); their properties are subsequently mapped onto the focal object. The resulting unconventional combinations are developed through free association, after which viable solutions are filtered. In his educational writings under the pseudonym G. Altov (2004), Altshuller provided a classic illustration: if the focal object is a pencil and the random object is a tiger, the emerging combinations include a "striped pencil," a "predatory pencil," or a "fanged pencil" (see Figure 1). Although TRIZ itself eventually moved beyond such stochastic approaches toward algorithmic problem-solving, this early reliance on semantic noise demonstrates how structured randomness can forcibly expand the associative field. Randomness works as a forced generator of semantic shifts, breaking the habitual functional fixedness of an object. Fritz Zwicky's morphological analysis proposes using a multidimensional matrix of parameters (a “morphological box”), where all possible variants of execution for different parts of the object are placed along the axes. A directed-random combination of values from different axes creates variants that logical enumeration would never yield. The matrix sets boundaries (contextual frame) while randomness fills the cells, saving the engineer from the dead end of “cliché recombination” (Zwicky, 1969).

Figure 1. The “predatory pencil” as interpreted by generative AI (NanoBanano)

In cognitive psychology, a large body of work shows that creative thinking involves not merely goal-directed search but two fundamentally different stages: a stochastic generation stage (bias-guided random proposal step), followed by a selective evaluation

stage (evaluation step depending on a flexible or transformable bias structure) (Simonton, 1999). In the philosophy of artificial intelligence, Margaret Boden proposed a taxonomy of creativity in which randomness occupies a clearly defined place. In her conception, combinatorial creativity – the generation of new combinations of known ideas – directly relies on stochastic mechanisms, and it is precisely "structured, not chaotic randomness that can generate originality." Boden distinguishes three types of creative processes: combinatorial (recombination of known elements), exploratory (movement within a given conceptual space), and transformational (change of the rules of the space itself). The first two of these directly depend on the system's ability to generate sufficient variety of options – an ability that, without a stochastic component, remains fundamentally limited.

GENERATIVE AI AS A COMBINATORIAL MACHINE

Despite outward similarity – a fixed vocabulary, combinatorics – modern transformers radically rethink the logic of combination. Instead of discrete symbols rotating on rigid axes, they use continuous vector representations, or embeddings. Each token is projected into a multidimensional latent space (dimensions on the order of several thousand), where semantic proximity is measured by cosine distance. Goodness and Greatness for Llull are isolated letters; in a transformer they are geometric neighbors, their correlation determined by the dot product of vectors. AI replaces the flat Llullian enumeration with a multidimensional continuous topology. The second principal difference is the attention mechanism (self-attention) which introduces dynamic dependency between tokens. Instead of fixed connections (as in Llull's disks), the transformer computes Query, Key, and Value matrices for each token, allowing each word to "weigh" all others depending on context (Luo et al., 2023). The token 'bank' shifts to entirely different coordinates within the vector space depending on whether its context includes 'river' or 'money.' The model computes not merely an isolated probability but a dynamic trajectory of meaning within the contextual field – a trajectory that is never fixed but recalculated for each new sequence.

However, even this architecture remains merely an averaging automaton, when it is operating in a deterministic selection mode where the model always chooses the most probable token – a state achieved when the "temperature" parameter approaches zero, stripping the system of any randomness. In terms of language, it transmits only the statistical norm, producing safe clichés. It transmits the statistical norm of language – safe clichés. Genuine interest, from both a philosophical and artistic standpoint, arises only when a stochastic component is introduced into the generation process.

The process of text generation in transformer models unfolds step by step as a sequence of mathematical transformations, each contributing to the final character of the utterance. The starting point is the computation of logits of unprocessed numerical scores that the model assigns to each token in its vocabulary. These values, not being probabilities, reflect the model's "confidence" in a given continuation. However, logits do not themselves form a distribution: their transformation into probability space is performed by the softmax function, which exponentiates (exponentially scales) the logits and normalises them, creating a smooth distribution where dominant values receive a

disproportionately large share of probability. It is precisely this “winner-takes-all” structure of softmax that creates the precondition for what the literature calls “neural text degeneration” – a situation where the model, following the most probable trajectory, produces predictable, template-like sequences (Finlayson et al., 2024).

Temperature (T) serves as a hyperparameter that scales the logits before applying softmax. To understand its mechanism, one must turn to the field of information thermodynamics, where the concept is borrowed from the Boltzmann distribution in statistical physics. In physical systems, higher temperature increases the thermal agitation of particles, allowing them to overcome energy barriers and move chaotically. In natural language processing, “temperature” acts as a mathematical proxy for this thermal noise, regulating the system's entropy or disorder. The behavior of the model scales precisely along this thermodynamic axis. At $T=1$, the distribution remains unchanged, representing the baseline linguistic state. At $T<1$, the distribution sharpens as the system “cools down”: the differences between high and low probabilities expand, freezing the model into a rigid, low-entropy state where it blindly selects the most probable tokens – a behavior that corresponds to strict mechanical determinism. Conversely, at $T>1$, the system "heats up" and the probability distribution flattens: peaks dissolve, entropy increases, and "tail" tokens – which normally carry high surprise value and almost never appear—gain substantially higher chances of being selected (Ficler & Goldberg, 2017). At extremely high temperatures, the distribution approaches uniform chaos, where token selection becomes virtually random.

This thermodynamic parameter is the technical correlate of what Borges (1983/1999) conceptualized as the subversion of order. In its “cold,” standard mode ($T \rightarrow 0$), the model strictly avoids combinations that exhibit low cosine similarity in latent space. For instance, the collocation “red tiger” is not linguistically impossible, yet because the model is trained on a statistical norm where “red” rarely modifies "tiger," this combination is suppressed. Only when the system is “heated” – allowing thermodynamic noise to shatter the rigid Lullian geometry of the linguistic matrix – can the model deviate from safe clichés and generate the “fanged” or “predatory” combinations that Altshuller observed in human creative association.

At elevated temperatures, the probability of selecting lower-ranked tokens increases, prompting the model to commit a statistical “error” by selecting an unlikely word. However, in generative systems, this error ceases to be a mere malfunction; instead, it actualizes hidden semantic connections dormant within the latent space as potential projections. The vector for “red,” for instance, extends beyond its overt color value to project onto concepts of “fire,” “fury,” “the mystical,” or “the forbidden.” When thermal noise forces the system to abandon the dominant statistical trajectory and select from the probability tail, it induces a sudden shift in contextual alignment. The semantic coordinates of the concepts converge at a single, previously unmapped point in the multidimensional topology, generating an emergent meaning: the “red tiger” ceases to be an statistical anomaly and becomes a powerful symbolic, demonic, or hallucinatory image. The transformer literally implements the Borgesian search for the ultimate epithet, doing so not through flat combinatorial enumeration, but through an instantaneous phase transition within its latent geometry. This is no longer a simple juxtaposition of two

lexemes, but a conceptual metaphor – one that was neither explicitly present in the training data nor programmed by the developer.

If the classical transformer architecture is in itself a "sighted" but deterministic system of contextual alignments, then genuine creativity and autonomy of code break into it through stochastic sampling. With temperature increasing to T = 1.5, the probabilistic envelope of the model instantly expands from one banal word to a horizon of 131 words. The top set of correct, predictable words loses its force – its mass "leaks" away by as much as 90%. The machine receives mathematical permission to step into the dark – to enter the "long tail" of unlikely, strange concepts. Another option for improving creative problem-solving is introducing noise at the input layer (seed-conditioning). Creativity is a process of restructuring the probabilistic field inside the "black box" at a moment of high incoherence (uncertainty) (Parupudi et al., 2026).

Many contemporary studies are devoted to how randomness and creativity can be increased without compromising truthfulness. West and Potts (2025), using tasks requiring unpredictable outputs – generation of random numbers, games with mixed strategies (rock-paper-scissors, hide-and-seek), and creative writing – demonstrate that "base models systematically beat aligned models at randomness and creativity." Alignment through reinforcement learning from human feedback (RLHF) narrows the behavioural repertoire of the model: it begins to prefer "7" over uniformly distributed numbers, becomes almost entirely predictable in certain game states, and sacrifices originality for "pleasant" writing. Moreover, the authors identify a stable correlation: the higher the model's performance on standard benchmarks, the worse its results on creativity tasks, indicating a fundamental trade-off between optimisation toward the average norm and capacity for deviation (West & Potts, 2025). Alignment narrows the probability distribution: RLHF models "concentrate probability on a set of anchor spans that co-occur across multiple generations for the same prompt." This leads to mode collapse – a significant drop in output diversity.

Traditionally, hallucinations of large language models are perceived as a defect – it amounts to the generation of false information with high confidence. However, if we view this phenomenon from a Borgesian perspective, we find that hallucination is the reverse side of the Llullian dream of infallibility. Modern models undergo reinforcement learning from human feedback, which encourages confident answers; leaderboards reward categoricalness. The model learns to "bluff": it issues a categorical assertion even when its internal state is characterised by high entropy – when vectors in the residual stream destructively interfere. Here a change is proposed: instead of attempting to suppress hallucinations, models should be shifted into a mode of calibrated uncertainty (Belrose et al., 2025). This means that the model should learn to correlate its linguistic confidence with its actual mathematical confidence, which can be computed, for example, from the entropy of the softmax distribution or the degree of conflict between vectors in the residual stream. If the model registers high entropy, it should signal this – using modal verbs, expressions of doubt, refusing categorical answers. In so doing, it ceases to pretend to be omniscient and becomes an agent that acknowledges the limits of its knowledge.

CONCLUSION

Ramon Llull sought to create a machine that would exclude error, doubt, and deviation. His *Ars Magna* was a project of totalitarian truth, where the rotation of disks replaced thought. Borges showed that this project is ontologically untenable: in a finite system of combinations there is no room for judgment, and therefore it generates only emptiness. Contemporary AI, especially in its artistic applications, has proven capable of realising precisely what Borges proposed as an alternative: using combinatorial redundancy not for proof but for generating new combinations, which then acquire meaning in the act of human interpretation.

As Borges convincingly demonstrated, this path inevitably leads to semantic collapse: a closed linear system, lacking contextual flexibility and criteria for internal verification, can only generate redundant syntactic noise, where any statement drowns in combinatorial explosion. This was the dream of an automaton that would exclude the very possibility of doubt, error, or deviation.

However, contemporary AI science paradoxically implements precisely the alternative that Borges proposed as a poetic antidote to the Llullian dead end. Generative neural networks overcome the blindness of medieval mechanisms by replacing discrete symbols with the continuous geometry of multidimensional semantic fields, and the rigid axes of disks with dynamic trajectories of contextual attention. But the crucial epistemic breakthrough occurs not at the level of increased computational power, but at the moment of mathematical legitimisation of randomness.

The introduction of stochastic sampling and the regulation of the temperature parameter have transformed AI from a deterministic calculator of linguistic clichés into a medium. When the model, pushed by temperature shift, takes a step into the “long tail” of improbable concepts, a phase transition occurs in the latent space. Contextual alignment collapses, bringing together previously incompatible semantic vectors at a single point.

Thus, the transformer, equipped with the capacity for stochastic deviation and the calibration of its own uncertainty, ceases to be a blind Llullian automaton and approaches the Borgesian ideal of a poetic machine – an instrument that does not assert but suggests; does not prove but hints; does not close truth but opens a space for play. It is precisely in this sense that Borges's diagnosis should be understood today: the thinking machine, freed from the obligation to be correct, can become productive precisely through its “incorrectness.”

REFERENCES

Altov, G. (2004). *Tvorchestvo kak tochnaya nauka: Teoriya resheniya izobretatelskikh zadach* [Creativity as an Exact Science: Theory of Inventive Problem Solving] (2nd ed.). Scandinavian Publishing House of the Official G.S. Altshuller Foundation.

Altshuller, G. (1996). *And suddenly the inventor appeared: TRIZ, the theory of inventive problem solving* (L. Shulyak, Trans.). Technical Innovation Center.

Belrose, N., Ostrovsky, I., McKinney, L., Furman, Z., Smith, L., Halawi, D., Biderman, S., & Steinhardt, J. (2025). Eliciting Latent Predictions from Transformers with the Tuned Lens. *arXiv*. https://doi.org/10.48550/arXiv.2303.08112

Borges, J. L. (1984). *Seven Nights*. (E. Weinberger, Trans). New Directions Publishing.

Borges, J. L. (1985). *Dreamtigers* (M. Boyer & A. Reid, Trans.). University of Texas Press. (Original work published 1960).

Borges, J. L. (1999). *Selected Non-Fictions*. (E. Allen, S. J. Levine, & E. Weinberger, Trans). Viking/Penguin Books.

Borges, J. L. (2005). *The book of imaginary beings* (N. T. di Giovanni, Trans.). Viking. (Original work published 1967).

Burroughs, W. S., & Gysin, B. (1978). *The Third Mind*. Viking Press.

Demina, A. I., & Nesterov, A. Y. (2025). *Semioticheskaya teoriya tvorchestva: Problema novogo v tekhnike i iskusstve* [Semiotic theory of creativity: The problem of the new in technology and art]. Mudraya Cherepakha.

Ficler, J., & Goldberg, Y. (2017). Controlling Linguistic Style Aspects in Neural Language Generation. In *Proceedings of the Workshop on Stylistic Variation*, 94–104. https://doi.org/10.18653/v1/W17-4912

Finlayson, M., Hewitt, J., Koller, A., Swayamdipta, S., & Sabharwal, A. (2024). Closing the Curious Case of Neural Text Degeneration. *International Conference on Learning Representations, 2024*, 12295–12313. ICLR.

Gogol, N. V. (1844). Gogol–Aksakov S. T. [Letter]. *Fundamental Electronic Library "Russian Literature and Folklore."*

Guízar, E. (2012). Borges y el delirante orden de las cosas en Ramón Llull. *Variaciones Borges*, *34*, 165–181.

Kircher, A. (1650). *Athanasii Kircheri... Musurgia universalis sive Ars magna consoni et dissoni in X. libros digesta*. Vol. 1-2. ex typographia haeredum Francisci Corbelletti.

Korsakov, S. N. (2009). *On the Application of Mechanics to Intellectual Operations*. MEPI. (Original work published 1832)

Kowalick, J. (1998, August 21). Psychological Inertia. *The TRIZ Journal*. https://the-trizjournal.com/psychological-inertia/

Lovelace, A. (1843). Notes by the translator to L. F. Menabrea's «Sketch of the Analytical Engine invented by Charles Babbage». *Taylor's Scientific Memoirs*, *3*, 666–731

Luo, Q., Zeng, W., Chen, M., Peng, G., Yuan, X., & Yin, Q. (2023). Self-Attention and Transformers: Driving the Evolution of Large Language Models. In *2023 IEEE 6th International Conference on Electronic Information and Communication Technology (ICEICT)*, 401–405. IEEE. https://doi.org/10.1109/ICEICT57916.2023.10245906

Lyubimov, Y. P. (2001). *Rasskazy starogo trepacha* [Tales of an Old Chatterer]. Novosti. (In Russian)

Mikhailov, A. S. (2016). Uspelenie vozmozhnostey razuma – izobreteniya S.N. Korsakova [Amplifying the Capacities of the Mind – S.N. Korsakov's Inventions]. *Artificial Intelligence and Decision Making*, *2*, 5–15.

Parupudi, V. S. R., Ponnada, H., Kaushal, A., Parupudi, S. S., Dasari, S., & Bulusu, S. (2026). Before and After Temperature: A Distributional View of Creative LLM Generation. *arXiv*. https://doi.org/10.48550/arXiv.2606.01451

Pliny the Elder. (1991). *Natural History*. [H. Rackham, Trans.] (Original work c. 77 CE)

Simonton, D. K. (1999). *Origins of Genius: Darwinian Perspectives on Creativity*. Oxford University Press.

Ugander, S., & Epstein, Z. (2024). Leonardo's Stains. *Journal of the History of Ideas*, *85*(3), 421–441.

Vroon, R. (1983). *Velimir Xlebnikov's Shorter Poems: A Key to the Coinages*. Department of Slavic Languages and Literatures, University of Michigan.

West, P., & Potts, C. (2025). Base Models Beat Aligned Models at Randomness and Creativity. *arXiv*. https://doi.org/10.48550/arXiv.2505.00047

Zhao, X., Zheng, B., Si, C., Yu, H., Liu, K., Zhou, R., Li, R., Chen, T., Li, X., Zhang, Y., & Wu, T. (2025). The Ramon Llull's Thinking Machine for Automated Ideation. *arXiv*. https://doi.org/10.48550/arXiv.2508.19200

Zwicky, F. (1969). *Discovery, Invention, Research Through the Morphological Approach*. Macmillan.

.

СВЕДЕНИЯ ОБ АВТОРЕ / THE AUTHOR

Дарья Быльева. bylieva_ds@spbstu.ru
ORCID 0000-0002-7956-4647

Daria Bylieva, bylieva_ds@spbstu.ru
ORCID 0000-0002-7956-4647

Статья поступила 3 марта 2026
одобрена после рецензирования 18 мая 2026
принята к публикации 10 июня 2026

Received: 3 March 2026
Revised: 18 May 2026
Accepted: 10 June 2026